Patient Level Analytics Using Self-Organising Maps: A Case Study on Type-1 Diabetes Self-care Survey Responses

Santosh Tirunagari*†, Norman Poh*, Kouros Aliabadi *, David Windridge†§ and Deborah Cooke ‡
* Department of Computing
† Center for Vision, Speech and Signal Processing
‡ School of Health Sciences
*†‡University of Surrey, Guildford, Surrey, United Kingdom GU2 7XH.
§Department of Computer Science, Middlesex University, The Burroughs, Hendon, London NW4 4BT.

s.tirunagari, n.poh, ka00115, d.windridge and d.cooke@surrey.ac.uk

***Abstract*—Survey questionnaires are often heterogeneous because they contain both quantitative (numeric) and qualitative (text) responses, as well as missing values. While traditional, model-based methods are commonly used by clinicians, we deploy Self Organizing Maps (SOM) as a means to visualise the data. In a survey study aiming at understanding the self-care behaviour of 611 patients with Type-1 Diabetes, we show that SOM can be used to (1) identify co-morbidities; (2) to link self-care factors that are dependent on each other; and (3) to visualise individual patient profiles; In evaluation with clinicians and experts in Type-1 Diabetes, the knowledge and insights extracted using SOM correspond well to clinical expectation. Furthermore, the output of SOM in the form of a U-matrix is found to offer an interesting alternative means of visualising patient profiles instead of a usual tabular form.**

I. Introduction

Knowledge and *information* are essential requirements of the present day world. Statistical machine learning algorithms can analyse data from different perspectives and summarise it to gain useful insights [1]. Such algorithms are also increasingly applied in understanding complex surveys related to human factors, health, bio-sciences [2] and social sciences [3] [4]. These complex questionnaires are mostly heterogeneous (with a variety forms of data, e.g. text, numerical, which exhibit missing values, wrong entries, imbalance and abnormalities), involving a large number of variables. Using traditional statistical methods such as null-hypothesis testing and descriptive techniques (mean, variance and frequency) might lead us to overly simplified conclusions [3]. When faced with such data, a combination of dimensionality reduction techniques [5] [6], unsupervised clustering [7] [8] and data visualisation techniques therefore have to be employed. There are many such methods available in the literature, including: Neighbour Retrieval Visualiser (NeRV) [9], t-Distributed Stochastic Neighbour Embedding (t-SNE) [10], Generative Topographic Mapping (GTM) [11] and Multi-Dimensional Scaling (MDS) [12] which could perform the aforementioned task-combination. In this study we have employed Self Organising Maps (SOM) [13] for survey data analytics because of its simplicity and also the faithfulness of its high-dimensional to low-dimensional mapping.

A. Motivation for using SOM

The process of analysing the data in the form of pictures is called information visualisation. This helps and supports decision making in numerous fields, including health-care surveys. Visualising information from large amounts of heterogeneous survey data in order to find out interesting patterns is a difficult task, but by using data-mining techniques (clustering) coupled with artificial neural networks in the form of SOMs renders it tractable.

In particular, clinicians often conduct surveys to better understand their patients. As mentioned earlier, using traditional descriptive statistical methods such as mean, variance, skewness and frequency, may lead to overly simplified conclusions. Hence, clinicians require statistical machine-learning tools that could be deployed as a 'black-box' for carrying out data analysis. For these reasons, we make use of the SOM algorithm for mining correlations and clustering similar responses within the surveys. The clustered responses in the higher dimensions are then visualised in a 2-dimensional grid thereby reducing the complexity within the data. Reducing the complexity in the data reveals more meaningful relationships, enabling understanding of the dependencies among the responses given in the survey. Previously, SOM has been used to visually explore data areas such as health, lifestyle, nutrition [14], financial [15], gene expression [2] [16], marine safety [17] and linguistics [18]. Recently, SOM has also been used to explore questionnaire based loneliness survey data [3].

B. Motivation underlying the current study

Type-1 diabetes is a major health problem in the present generation with 10% of all the adults are diagnosed with diabetes. There are many factors that must be considered to effectively manage it; daily insulin injections, a healthy diet, regular physical activity as well as others described later in Table III. Type-1 diabetes can develop at any age but usually appears before the age of 40. It is the most common type of diabetes found in children [19]. It is often influenced by the

lifestyle of the patient and treatment requires well managed self-care. Awareness of medicine adherence also plays an important role in the treatment. We therefore, in this study, aim to locate, define, analyse and interpret, via statistical machine learning approaches, patterns existing in the habits and behaviour of patients with regard to their medication in order to motivate treatment suggestions and determine the most suitable treatment plan.

C. Objectives

The main objectives of this study are two-fold. First, from the computational perspective, we would like to examine the feasibility of using Self-Organizing Map as a means of extracting useful information from survey questionnaires. Second, from the scientific perspective, we would like to understand if the responses collected from the Type-1 Diabetes survey are reasonable and correspond to what domain experts and clinicians would expect. For instance, it is desirable to answer the following questions of an exploratory nature:

- Can we identify co-morbidities from the survey?
- What are the self-care factors or behaviours that are dependent on each other?

The questions being posed here cannot be readily answered using classical methods such as generalized linear models and their variants; hence, the motivation for using SOM for exploring the data.

D. Contributions

Our contributions can thus be summarised as follows:

- **Novel use of SOM for visualising individual patient data** Although SOM has been widely used, its uses in visualising individual patient profiles are rarely highlighted or discussed. Our approach of summarising and visualising individual patient profiles turns will prove to be useful in this, as concurred by the domain experts.
- **Improved understanding** We will demonstrate that the visual analytics provided by SOM can improve experts' understanding of the impact of self-care behaviours of patients with Type-1 diabetes. (For instance, we establish that factors relating to food consumption behaviours are closely clustered within the SOM, as are factors relating to insulin management).

We therefore demonstrate that SOM is a potentially viable tool for analysing high-dimensional questionnaire responses as well as a means for visually summarising individual patient data.

E. Organisation

The organisation of the paper is as follows: In section II, we present and analyse the survey dataset and illustrate the demographics of the data. In section III, we present the SOM methodology used in this paper. Experiments and results are discussed in section IV including data preprocessing and imputation (filling out the missing values). Finally, in section V, we draw conclusions and summarise the discussions.

II. Dataset, Preprocessing & Demographics

The survey consists of 611 patients' responses (all above 18 years old and with Type-1 diabetes), which includes 15 questions on self-care factors. The questionnaire also took responses for the co-morbidities associated with the Type-1 diabetes. The responses for these co-morbidities are a binary 'Yes' or 'No'. The responses for the self-care behaviours are required to be one of the following: 1) Never 2) Rarely 3) Sometimes 4) Usually and 5) Always. The data was collected not only through the medium of paper but also on-line. It contains both unstructured text and numerical data. The abnormalities in the data are the missing values and wrong entries (anomaly or outliers).

A. Data Preprocessing

The processing of the data included the conversion of string values (yes, no, always, often, sometimes, rarely, never, missing (flagged as Not-a-number, NaN) and Not Applicable (NA) into categorical numerical values for ease of computation. This conversion makes the data analysis computationally inexpensive. For 'NA' results we used 0, always, often, sometimes, rarely and never are scored as 5, 4, 3, 2 and 1 respectively, and for NaNs the appropriate missing value. K-nn base imputation (via Matlab's '*knnimpute(Data)*') was used to replace NaNs in the data with the corresponding value from the nearest-neighbour column. The nearest-neighbour is the closest neighbour in Euclidean distance terms. If the corresponding value from the nearest-neighbour column-vector is also NaN, the next nearest column-vector is used.

TABLE I. Table showing the skewness within the data (problem of veracity).

Ethnicity	**Count**	**Percent**
White British	558	91.33%
Other White	32	5.24%
Mixed	2	0.33%
British Asian	3	0.49%
Black British	4	0.65%
NA	4	0.65%
Other Ethnic Group	3	0.49%
Asian	3	0.49%
Black	2	0.33%

Marital	**Count**	**Percent**
In a significant relationship	467	76.43%
Single	95	15.55%
Divorced / Separated	36	5.89%
NA	5	0.82%
Widowed	8	1.31%

Employer	**Count**	**Percent**
Self-employed / Freelance without employees	57	9.33%
Employee	515	84.29%
Self-employed with employees	20	3.27%
NA	19	3.11%

The Table I refers to the veracity problem where the data is imbalanced towards one variable. For instance, 'white British' constitute 91.33% of the cohort.

B. Demographics

We apply descriptive statistical methods for analysing the demographics in this survey. Participants in the survey are required to be at least eighteen years old and the eldest participants in the survey are over eighty. The highest percentage of

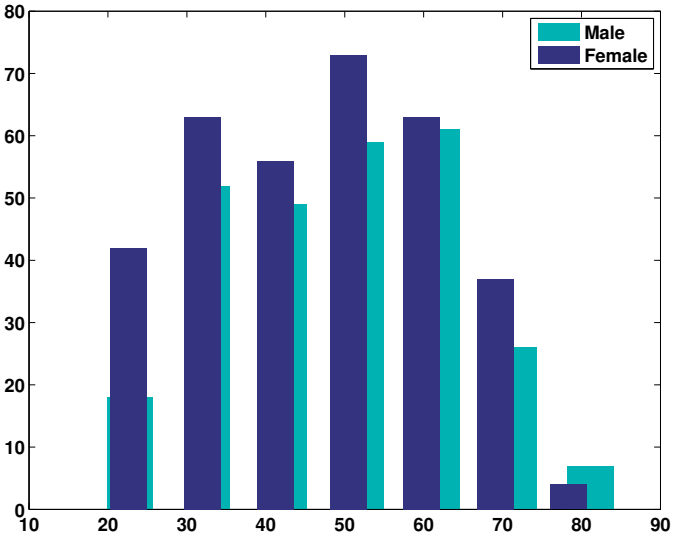

Fig. 1. Distribution of age with the majority of participants being between the ages of 30 and 60.

respondents are in their mid-thirties to early sixties as shown in the histogram in Figure 1.

Male participants represent 45% of the survey population while female respondents constitute 55%. The most common employment status is full time employment at 48.45% and the lowest percentage being unemployed is 3.11%. Modern professional occupation is the dominant profession within the sample data and contribute 23.4% of the respondents but clerical and intermediate occupations also contribute 19.64%.

TABLE II. TABLE SHOWING THE DEMOGRAPHICS ASSOCIATED WITH HYPO.

Hypo levels	**Count**	**Percent**
less than 3.0 mmol/L	267	43.70%
greater than or equal to 3.0mmol/L	317	51.88%
do not feel symptoms	22	3.60%
NA	5	0.82%

Hypo awareness	**Count**	**Percent**
Hypo unaware	289	47.30%
Hypo aware	317	51.88%
NA	5	0.82%

The majority of the patients in the survey data responded that the symptoms of hypoglycemia occurred at blood glucose levels of greater than or equal to 3.0mmol/L (see, Table II). The percentage of respondents that are aware of hypos commencing was over half but only by a small margin, with 51.88% of patients being aware.

The most common co-morbidity in the dataset is Retinopathy followed by high BP and high cholesterol (see, Figure 2). The most common complication found in females who have Type-1 diabetes is Polycystic ovary syndrome and in males is sexual dysfunction. Haemoglobin levels for patients who have Type-1 diabetes are generally higher in males.

We are interested in alternative means of visualising patient profiles instead of a usual tabular forms and histograms. For this purpose, we study the existing Type-1 diabetic patient behaviours based on their self caring factors shown in Table III using SOM method, which we shall discuss in the next section.

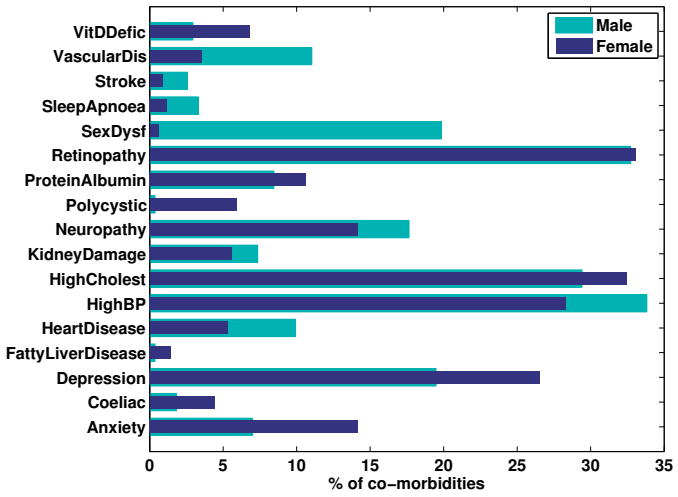

Fig. 2. Bar-chart showing the percentage of co-morbidities associated with Type-1 diabetes.

TABLE III. SELF-CARE FACTORS PRESENTED IN THE SURVEY.

Label	**Self-care factors**
CBG-Monitor	Check blood glucose with monitor
RBG-Results	Record blood glucose results
CKGL-High	Check ketones when glucose level is high
TCD-Insulin	Take correct dose of insulin
TI-Time	Take insulin at the right time
EC-Food Portions	Eat the correct food portions
Eat-Timely	Eat meals/snacks on time
KF-Records	Keep food records
RF-Labels	Read food labels
Rec-Carbs	Treat low blood glucose with just the recommended amount of carbohydrate
Carry-Sugar	Carry quick acting sugar to treat low blood glucose
Clin-Appoint	Come in for clinic appointments
WM-Alert	Wear a medic alert ID
Excercise	Exercise
AIDGFE	Adjust insulin dosage based on glucose values, food, and exercise.

We are also interested in studying the correlations among the co-morbidities associated with Type-1 diabetes. The questionnaire is presented with the co-morbidities shown in Table IV.

TABLE IV. CO-MORBIDITIES ASSOCIATED WITH TYPE-1 DIABETES.

Anxiety	Heart disease / heart attack	Coeliac disease
High blood pressure (hypertension)	Depression	High cholesterol (triglycerides / lipids)
Fatty liver disease	Kidney damage / renal failure	Neuropathy (damage to the nerves in feet)
Polycystic ovary syndrome (women only)	Retinopathy (damage to the eye (retinal))	Protein (albumin) in the urine
Sleep apnoea	Sexual dysfunction	Stroke
Vitamin D deficiency	Vascular disease (poor circulation in legs / feet)	

III. METHODS

To explore the survey dataset we have used Self organising map (SOM). SOM also known as Kohonen map [13] is an unsupervised technique that is most often described in the language of artificial neural networks. SOM provides a way of representing multidimensional data in typically two or three dimensions. This process of reducing the dimensionality of

vectors is based on a data compression technique known as vector quantisation. In addition, SOM creates a network that stores information in such a way that any topological relationships within the training set are maintained. Hence, SOMs are useful for visualising large data sets of high dimensionality. SOM is an unsupervised, competitive learning approach in which only one neuron 'wins' each training phase. There are no connections between the neurons in the input and output layers. However, they communicate with each-other via a neighbourhood function. If a neuron wins during the training phase, it will also impact its neighbours.

Let us consider the input vector

$$\boldsymbol{x} = [x_1, x_2, ..., x_n]^T \tag{1}$$

The synaptic weight vector of the neuron i in the output layer of 2-dimensional neurons is

$$\boldsymbol{w_i} = [w_i^1, w_i^2, ..., w_i^n]^T, i = 1, 2, \cdots, m; \tag{2}$$

where m is the number of output neurons and w_i^v is the weight associated with neuron m and variable v. Although the output neurons are arranged in 2-dimensional array, their weight vectors are n-dimensional i.e. the same dimensions as the input vector $\boldsymbol{x}$. The Euclidean distance $\|\boldsymbol{x} - \boldsymbol{w_i}\|_2$ of the current input vector $\boldsymbol{x}$ to all of the weight vectors $i = 1, 2...m$ is computed. The winning neuron is one whose weight vector $\boldsymbol{w_q}$ has the minimum Euclidean distance to $\boldsymbol{x}$, i.e.,

$$q(x) = arg_{min_i} \|\boldsymbol{x} - \boldsymbol{w_i}\|_2 \tag{3}$$

The weight vectors of the winning neurons and the neurons in its predefined neighbourhood η_q are updated using gradient descent, leading to the following update rule:

$$w_i(k+1) = w_i(k) + \eta_{qi}(k)[x(k) - w_i(k)] \tag{4}$$

Neurons outside the neighbourhood are not updated i.e $\eta_q(k) = 0$ and neurons inside the neighbourhood η_q are updated using equation 5:

$$\eta_{qi}(k) = \mu(k). \tag{5}$$

The learning parameter $\mu(k)$, where $0 < \mu(k) < 1$ decreases with increasing iterations. The learning process has two phases: 1) ordering phase (rough training phase) and 2) convergence phase (fine training phase). In the ordering phase, topological ordering of the weight vectors is carried out. The learning parameter $\mu(k)$ is set close to unity. In the convergence phase, the self-organising map is fine-tuned, which is achieved by setting the learning parameters $\mu(k)$ to the order of 0.01. The stopping criterion for the SOM algorithm is the number of specified iterations, or else a sufficiently small degree of change in the weight vectors.

IV. Experiments and Results

As mentioned earlier, traditional descriptive statistical methods such as mean, variance, skewness and frequency, may give clinicians overly simplified conclusions for their surveys. Hence, to obtain a deeper level understanding of these surveys, and to better understand self-care behaviours for each individual patient, we apply a SOM to depict patient level analytics. We hence, in this section, conduct three experiments as follows:

- Determination of the correlations among co-morbidities associated with Type-1 diabetes.
- Identification of patient profiles associated with co-morbidity.
- Identification of patient profiles based on their self-care behaviours.

A. Correlations amongst co-morbidities

Patients with Type-1 diabetes often suffer with other diseases which effect their self-care behaviours. This may result in insulin resistance. Thus, it is necessary to determine significant correlations existing among co-morbidities. To achieve this, we introduce the 611 patient responses with their 17 co-morbidities ($611 - by - 17$) to the SOM algorithm[1]. The SOM outputs a text visualisation map of co-morbidities and its visualisations of cross-correlations. In this experiment we have chosen a 30×20 grid for visualisation.

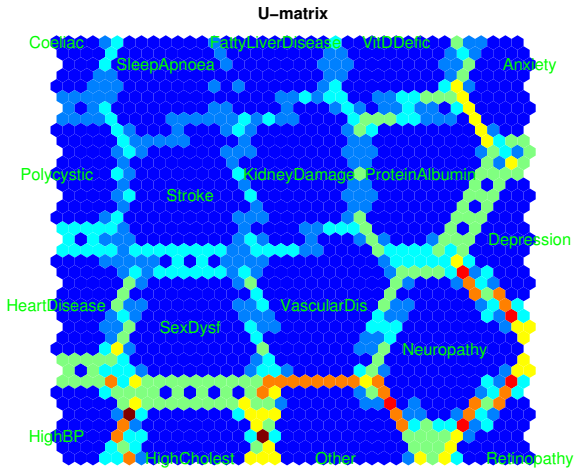

Fig. 3. Unified distance matrix (U-matrix) showing different clusters on the grid. Blue colour represents clusters whereas other colours can be considered as the cluster separators.

The clusters in Figure 3 depict the distances between the neurons using different colours between the adjacent nodes. A blue colour is indicative of clusters whereas the other colours are the cluster separators. Results are achieved by unsupervised learning, that is, without human intervention. Representing a SOM via the U-matrix thus offers an intuitively appealing way to gain insight into the data distribution [20].

The component planes could be visualised as cut planes or slices of the U-matrix. By comparing component planes one can see whether two components correlate or not. If the outlook is similar, the components strongly correlate. For example, in Figure 4, high BP and high cholesterol correlate with each other. Hence, the bottom left side of the U-matrix in Figure 3 reveals that high BP and high cholesterol have been clustered nearby. Similarly, the correlated co-morbidities (see Figure 4) are clustered nearby in the U-matrix (Figure 3). For example, we observe the following natural clustering of variables: (1) high BP and high cholesterol; (2) anxiety and depression; (3) heart disease and vascular disease; and, (4) Kidney damage and protein albumin.

[1] http://www.cis.hut.fi/projects/somtoolbox/

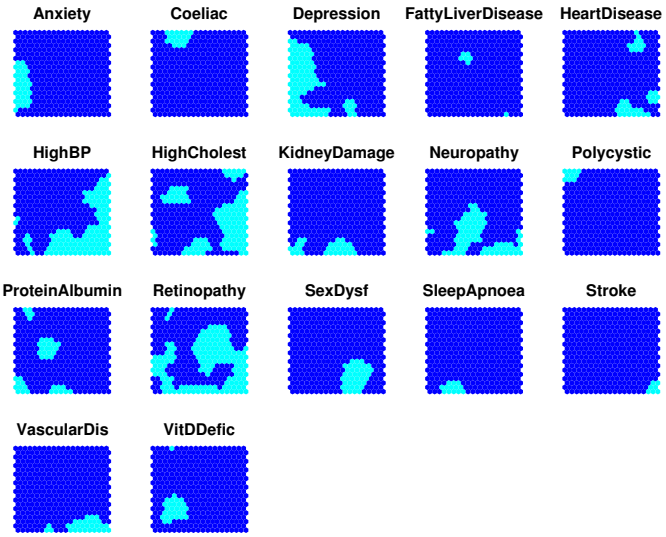

Fig. 4. Component planes representing the relative component distributions of the input data. Light blue values represent patients with the corresponding co-morbidity.

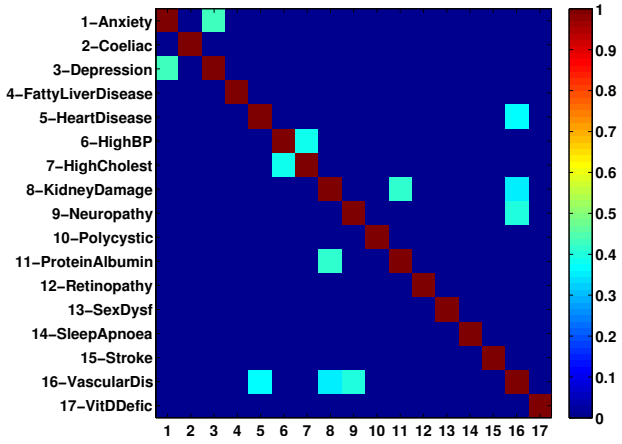

Fig. 5. Matrix showing correlations which are greater than 0.3 amongst the co-morbidities.

We are interested to see how these correlations in general are with in the data. For this, we have computed the correlations on the raw data-matrix ($611-by-17$) with values that are greater than 0.3, discarding the weaker correlations (less than 0.3). The results of our correlation analysis in Figure 5 and the results in the U-matrix of SOM (Figure 3) strongly agree, as we find stronger correlations between (1) high BP and high cholesterol; (2) anxiety and depression; (3) heart disease and vascular disease; and, (4) Kidney damage and protein albumin.

B. Patient level analytics on co-morbidities

Clinicians in general are interested in co-morbidities in their patients and how this affects the treatment plan. To achieve this, we introduced Patient ids (N=611) and their co-morbidities (d=17) as an input (matrix of size $18-by-611$) to the SOM. This resulted in 611 component planes, revealing individual patients and their dependencies/co-morbidities.

The component planes in Figure 6 indicate four patient ids (corresponding to $\{111, 211, 311, 611\}$) along with their

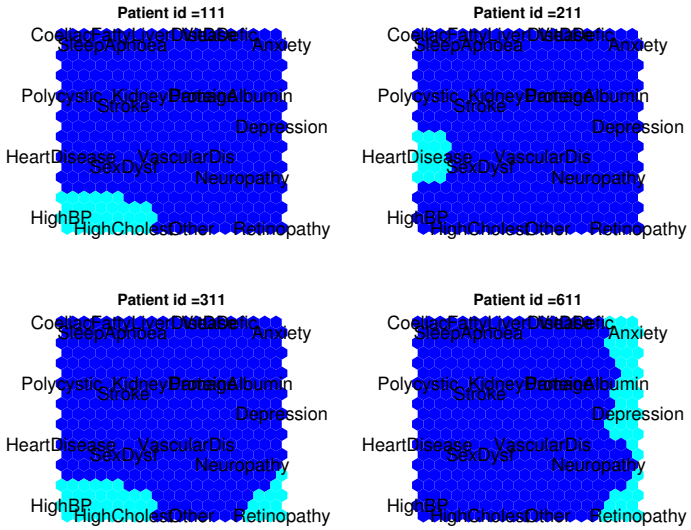

Fig. 6. Component planes of patient ids = $\{111, 211, 311, 611\}$ showing their associations with the co-morbidities.

associated co-morbidities. Patient id = '111' is associated with high BP and cholesterol, patient id = '211' with heart disease, patient id = '311' with high BP, high cholesterol and Retinopathy and patient id = '611' with the depression, anxiety and Retinopathy. A component plane of a single patient associated with no co-morbidities is displays a plain image by way of reference.

C. Patient level analytics on self-care factors

Improving the self-care behaviours can improve quality of life. Hence, it is important for a clinician to know how well their patients are taking care of themselves. This helps clinicians to suggest suitable treatment plans in accordance with the life-style of their patients. To visualise how self-caring the patients are, we have introduced the self-care factors ($d = 15$) described in section I-B along with the patient ids ($N = 611$) (matrix of size $15 - by - 611$) to the SOM. The output showed 611 component planes revealing individual patients and their associated self-care factors.

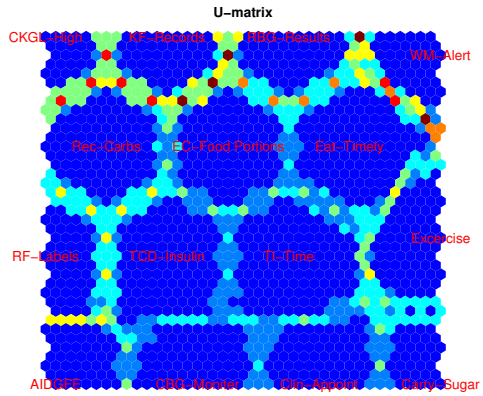

Fig. 7. U-matrix showing the clustering results of the self-care factors.

The U-matrix in Figure 7 shows the clustering results of the self-care factors. The U-matrix is clearly divided into 15

clusters, each representing a self-care factor. Factors relating to food are clustered closely as are factors relating to taking insulin.

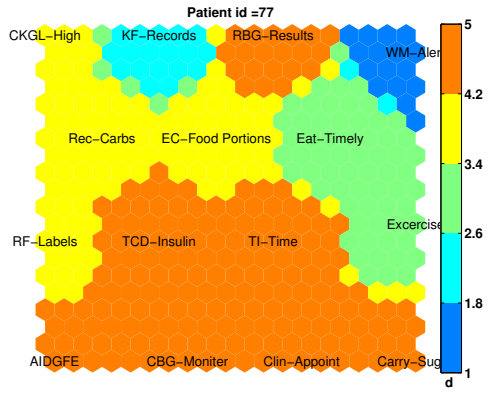

Fig. 8. Component plane of patient id corresponding to 77 showing the associations with the self-care factors.

The component plane in Figure 8 shows patient id 77 and his/her dependencies with respect to the self-care factors. The colorbar in this figure shows the user rating i.e 'Never' corresponding to darker blue, 'Rarely' to lighter blue, 'Sometimes' to green, 'Usually' to yellow and 'Always' to orange. It is apparent from Figure 8 that patient id = 77 is not wearing a medical alert and rarely keeps food records.

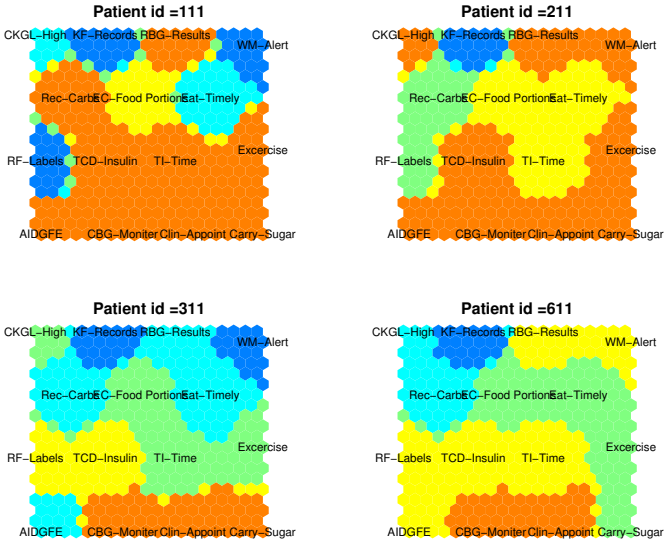

Fig. 9. Patient ids = $\{111, 211, 311, 611\}$ and their dependencies with regard the self-care factors.

Similarly, the component planes in Figure 9 show four patient ids = $\{111, 211, 311, 611\}$ and their dependencies with respect to the self-care factors. It is seen that none of the considered patients keep food records and all of them keep clinical appointments and check blood glucose levels via the monitor. These results are potentially of great interest for clinicians in understanding their patients' self-caring behaviours.

V. Conclusions and Discussion

In this study, we show the potential of using Self Organizing Maps (SOMs) as a statistical machine learning method for analysing survey data. In a survey study aimed at understanding the self-care behaviour of 611 patients with Type-1 Diabetes, we demonstrated that SOMs can be used to (1) identify co-morbidities; (2) to link self-care factors that are dependent on each other; (3) to visualise individual patient profiles; Although SOMs have been previously used to process survey data before [3], the use of SOMs for representing and visualising individual patient profiles, as well as for clustering patients is novel. Both usages turn out to be clinically useful, as concurred by clinicians and domain experts, because SOM can provide a visual summary of individual patient profiles, allowing them to group similar patients together.

Acknowledgement

The funding for this work has been provided by Department of Computing and Centre for Vision, Speech and Signal Processing (CVSSP) - University of Surrey.

References

[1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI magazine*, vol. 17, no. 3, p. 37, 1996.

[2] P. Törönen, M. Kolehmainen, G. Wong, and E. Castrén, "Analysis of gene expression data using self-organizing maps," *FEBS letters*, vol. 451, no. 2, pp. 142–146, 1999.

[3] K. Lagus, J. Saari, I. T. Nieminen, and T. Honkela, "Exploration of loneliness questionnaires using the self-organising map," in *Artificial Neural Networks and Machine Learning–ICANN 2013*. Springer, 2013, pp. 405–411.

[4] B. Castellani and F. W. Hafferty, *Sociology and complexity science: a new field of inquiry*. Springer, 2009.

[5] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse processes*, vol. 25, no. 2-3, pp. 259–284, 1998.

[6] F. Fallucchi and F. M. Zanzotto, "Singular value decomposition for feature selection in taxonomy learning," in *Proceedings of the International Conference RANLP-2009*. Association for Computational Linguistics, 2009, pp. 82–87.

[7] M.-S. Paukkeri, I. Kivimäki, S. Tirunagari, E. Oja, and T. Honkela, "Effect of dimensionality reduction on different distance measures in document clustering," in *Neural Information Processing*. Springer, 2011, pp. 167–176.

[8] S. Tirunagari, M. Hanninen, A. Guggilla, K. Stahlberg, and P. Kujala, "Impact of similarity measures on causal relation based feature selection method for clustering maritime accident reports," *Journal of Global Research in Computer Science*, vol. 3, no. 8, pp. 46–50, 2012.

[9] J. Venna, J. Peltonen, K. Nybo, H. Aidos, and S. Kaski, "Information retrieval perspective to nonlinear dimensionality reduction for data visualization," *The Journal of Machine Learning Research*, vol. 11, pp. 451–490, 2010.

[10] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of Machine Learning Research*, vol. 9, no. 11, 2008.

[11] C. M. Bishop, M. Svensén, and C. K. Williams, "Gtm: The generative topographic mapping," *Neural computation*, vol. 10, no. 1, pp. 215–234, 1998.

[12] J. B. Kruskal, "Nonmetric multidimensional scaling: a numerical method," *Psychometrika*, vol. 29, no. 2, pp. 115–129, 1964.

[13] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[14] Y. Mehmood, M. Abbas, X. Chen, and T. Honkela, "Self-organizing maps of nutrition, lifestyle and health situation in the world," in *Advances in Self-Organizing Maps*. Springer, 2011, pp. 160–167.

[15] G. Deboeck and T. Kohonen, *Visual explorations in finance: with self-organizing maps*. Springer London, 1998, vol. 2.

[16] P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. S. Lander, and T. R. Golub, "Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation," *Proceedings of the National Academy of Sciences*, vol. 96, no. 6, pp. 2907–2912, 1999.

[17] S. Tirunagari, M. Hanninen, K. Stanhlberg, and P. Kujala, "Mining causal relations and concepts in maritime accidents investigation reports," *International Journal of Innovative Research and Development*, vol. 1, no. 10, pp. 548–566, 2012.

[18] T. Honkela, S. Kaski, K. Lagus, and T. Kohonen, "Websomself-organizing maps of document collections," in *Proceedings of WSOM*, vol. 97, 1997, pp. 4–6.

[19] T. M. Frayling, N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy, C. M. Lindgren, J. R. Perry, K. S. Elliott, H. Lango, N. W. Rayner *et al.*, "A common variant in the fto gene is associated with body mass index and predisposes to childhood and adult obesity," *Science*, vol. 316, no. 5826, pp. 889–894, 2007.

[20] A. Ultsch and H. P. Siemon, "Kohonen's self organizing feature maps for exploratory data analysis," in *Proc. INNC'90, Int. Neural Network Conf.* Dordrecht, Netherlands: Kluwer, 1990, pp. 305–308.

[21] G. Soltesz, C. Patterson, and G. Dahlquist, "Diabetes in the young: a global perspective," *IDF Diabetes Atlas. Brussels: International Diabetes Federation*, 2009.